METHODS IN MAKING A CONCORDANCE

The editors of the concordance to the Latin works of Dante have come to the end of their task with a strong belief in the value of concordances, an entire willingness to leave to others the carrying on of the good work, and a certain stock of opinions as to methods in concordance-making which it seems desirable to record here for the possible benefit of future laborers.

With regard to the functions proper to a concordance we are entirely in agreement with the opinions expressed by Professor McKenzie in a recent report of this society:[1] a concordance may enable one who has in mind part or all of a phrase to ascertain readily the whole phrase and its location; it may serve as a register of the author's linguistic usage; and it may provide a full index of the subjects treated in the work or works covered. The value of the second function is illustrated in the accompanying article by Professor Rand. The value of the third may be instanced by the control of Dante's ideas as to the freedom of the will afforded by the series of articles *liber* to *libertate* in the three Dante concordances.

Among the problems best decided before the actual beginning of registration is that as to whether the several forms of an inflected word shall be presented in different groups or under a single heading: for example, whether the forms *aberat, abesse, absit,* and *absum* shall have each a separate article, or be grouped all under *absum.* Professor Fay in the *Concordance of the Divina Commedia* followed the first method; Professor Sheldon, in the concordance to Dante's minor Italian works, the second. We hold strongly with Professor Sheldon and Professor McKenzie[2] for the second method. It is an open question whether the first or second method better serves the first of the three concordance functions, and the second and third functions are much better served by the second method.

[1] *Means and End in making a Concordance,* in the *Twenty-fifth Annual Report of the Dante Society* (for 1906), p. 26.

[2] See pp. 35–40 of the article referred to.

Most words in a concordance have a line of context given for each occurrence. Such treatment is, however, obviously inappropriate for the common pronouns, prepositions, conjunctions, and certain other particles. The question as to what words shall receive less than complete treatment, and as to the extent and manner of treatment to be accorded them, is a difficult one. Full context should be given for all words which a user of the concordance might select as means of identification of a given passage, and for all words of any rarity in the work or works treated. For words of merely linguistic interest, in general, a complete series of references without context should be given. Such words, however, often assume a special importance, as for example *quia* in " State contenti, umana gente, al *quia* " (*Purgatorio* iii. 37), or in " Scientes *quia* rationale animal homo est " (*De vulgari eloquentia* ii. 10. 1). Such occurrences should have full context. To the occurrences thus allowed full context may well be added a few others typical of the author's general use of the word. Other occurrences of distinct but minor linguistic interest, for which brief context is sufficient, may well be registered in parentheses following the line references. This treatment is especially appropriate for favorite combinations of particles and for brief stock phrases. A series of articles containing such parentheses will afford intimate acquaintance with the author's linguistic individuality. Words of this type should therefore have in the concordance two divisions: a first division consisting of a series of occurrences with full context, and a second division consisting of a series of line references, some of which are followed by parentheses. A very few words, finally, are of such frequent occurrence that a student investigating their usage would find a list of references of no advantage. For these words occurrences of special importance should be registered, and the others disregarded. The only words so treated in our concordance are *et*, the relative *qui*, and *sum*. An initial decision as to which words shall receive which treatment can hardly be more than tentative. It should err rather by assigning too many than too few to the class not receiving full context, and rather by assigning too few than too many to the class of those to be in general disregarded.

The propriety of according treatment to direct quotations appearing in the work or works in question has been doubted. We feel that they

should be treated. Quotations in indirect discourse or in partial para-
phrase are necessarily treated; it seems illogical not to extend the same
treatment to direct quotations. Moreover, a user of the concordance
might very possibly desire to determine the form or location of a quota-
tion, and the content of a quotation often adds to our knowledge of the
author's knowledge of a given subject, or illuminates his linguistic usage.
In the registration of words occurring in a direct quotation the fact of
such occurrence should in some way be noted. In the case of words
having full context, the quotation marks should be retained as in the
text. In the case of words for which references only are given, the line
reference in question may be included in quotation marks.

The cards for use in the preparation of a concordance should be plain
rather than ruled, and about six inches by two inches in size. Before
the actual work of registration is begun, a rough estimate of the number
of words in each chapter (or equivalent subdivision) should be made;
then for each chapter a number of cards equal to the estimated number
of words in that chapter should be stamped in the lower left-hand corner
with a reference consisting of the abbreviated name of the work (if more
than one work is treated), the book or part, and the chapter: for example,
for the first chapter of the *De vulgari eloquentia* 200 cards should be
stamped *V.E.* i. 1. This work, of course, may be done by an assistant
without special knowledge.

Next, each word in the text (except the very few that are to be dis-
regarded) should be registered on a card by writing in the upper left-
hand corner the word in question in its proper index form (nominative
singular, present infinitive, or present indicative first singular, as the case
may be), and adding to the stamped book and chapter reference in the
lower left-hand corner the proper line reference. This work may be
done by an assistant who writes a clear hand and has enough linguistic
knowledge to reduce the words to the proper index form. The accuracy
and completeness of the work thus done should then be verified. It will
prove more economical, both in time and in mental energy, to verify first
the index forms on the whole series of cards and then the line references
on the whole series of cards, than to verify card by card first with regard
to the index form and second with regard to the line reference. This
verification should be done by two persons, one reading from the text,
the other handling the cards.

The next step should be, in our opinion, the insertion of the context upon all cards except those for words to which it has been tentatively decided to accord less than complete treatment. The writing of the contexts while the cards are in textual order rather than after alphabetization makes it possible to deal once for all with particular difficulties in thought and in syntax, and facilitates consistency in the treatment of similarly conditioned words.

The context for a single occurrence should not be so long as to run over one line in the printed concordance. A maximum length must be decided upon: with us it was 64 letters, — roughly, a line and a half of the Oxford Dante. It is desirable that all contexts should approach fairly closely the maximum length, since, if they do, the use of leaders on the printed page will be unnecessary, and the right-hand edge of the column of contexts will be agreeably even. In case a word occurs twice within a passage which is twice the length of the standard context passage, we think it wise, in most cases, to make the context for the two occurrences continuous, indenting the second line of it, rather than to divide the passage into two separate contexts.

In the writing of contexts, the use of omission dots is often necessary, but should be limited as strictly as possible, since the user of the concordance, if confronted with a context containing omission dots, is likely to have the fear that the portion of text omitted may contain something which concerns him. It will save much space to denote omission not by three dots but by two, as in the Oxford English Dictionary and in Mr. Rayner Storr's *Concordance to the De Imitatione Christi.*[1]

We think it unwise to rely upon assistants, however competent, to do the original writing of the contexts. Before the cards for our concordance were given into our hands, nearly all the contexts had been written out, some fifty assistants sharing the work. With all due appreciation of the generosity with which this work was done, we judge that it cost us more time and energy to correct, standardize, and revise the work thus presented, than it would have taken to do the work ourselves in the first place. If assistants are to be employed at all for the writing of contexts, it should be only after the editors have done enough of the work to have established certain well-defined principles of context selection; and even so, the assistant should work under the direct supervision of an editor.

[1] Oxford, 1910.

If principles of context selection change considerably in the course of the work, as is very likely to be the case, a revision of the contexts should follow immediately upon the completion of their original registration.

In the alphabetization, colored cards with projecting caps should be used to head the several articles, each colored card having written upon the cap the word in question in its index form.

Immediately after alphabetization, the words to which it was tentatively decided to accord less than complete treatment should be studied; final decision should be reached as to which of these words, if any, should be accorded full treatment; and, in the case of the remaining words, decision should be reached as to which occurrences should be accorded full context, which accorded parenthetical treatment, and which indicated by reference only.

The concordance should then be typewritten on the largest available size of typewriter paper. This expense we think fully justified: it insures the existence of the concordance in duplicate, it affords a clear copy for the typesetter, and it enables the editors to standardize the several articles and to anticipate the look of the printed page far better than can be done with cards.

The first context of each article should be typewritten upon the same line as the caption of the article, since this arrangement is necessary on the printed page in order to avoid expensive waste of space. In case the combined length of a caption and an initial context is so great as to make the context run over the line, the length of the context should be reduced in the final revision referred to below.

The typewriting once done, a careful verification of its accuracy should follow, one person reading from the cards and another following the typewriting, or vice versa. This work can be done by competent assistants.

There should then follow a final revision of the concordance, with the particular object of attaining consistency in treatment within each article.

E. H. WILKINS